On Commercialization of Scandinavian Machine Translation

Torbjørn NORDGÅRD
Dept of Language and Communication studies,
Norwegian University of Science and Technology (NTNU)
NO-7492 Trondheim
Norway
torbjorn@hf.ntnu.no

Tronn SKJERSTAD
LingIT AS
Elgesetergt 10
NO-7030 Trondheim
Norway
tronn@skjerstad.net

Abstract

This paper discusses basic issues connected to commercialization of academic machine translation prototypes aimed at the Scandinavian market. We maintain that Technology Transfer Offices at the universities are important tools for the researchers, and two concrete business models are discussed in some detail. We conclude that the most realistic commercialization strategy for an academic Scandinavian MT prototype is through creation of a new language technology company, alternatively allowing an existing software company the use of the prototype more or less for free. Even though two Norwegian prototypes are used as examples in this paper, we believe that the reflections are valid for other Scandinavian countries as well.

1   Introduction

Many machine translation system prototypes have been made over the last decades. But very few have reached the market or even approached the market via serious commercialization initiatives. In this paper we will focus on the path from MT prototype to marketable product. Our remarks are based on general knowledge of commercialization initiatives, and practical experience with commercialization of authoring tools containing spell-checkers, word prediction, dictionary access and text-to-speech interfaces, among other things via LingIT AS in Trondheim, Norway. LingIT (www.lingit.no) is currently working with a bilingual authoring tool where MT is absent, at least for the time being.

We will use Systran as a reference example of a successful MT company. Systran has revenues of 9,2 million EUR in 2001 (http://www.systransoft.com/Investors/AnnualRep ort2001-EN.pdf). Licenses are sold for $59 or 50 EUR on the private consumer market. Roughly, that should mean approximately 180000 licenses if all sales were private licenses (which is not the case since Systran operates in many market segments, but we make the assumption anyway). Given that there are 36 language pairs in the portfolio, we are left with approximately 5000 licenses per language pair (in one direction) on average. The language societies involved are bigger than what we find in Scandinavia, so the average 5000 licenses a year is a somewhat optimistic figure for the volume of an imagined Systran system between Norwegian and English. Let us assume 4000 licenses a year, which gives a very rough turnover estimate of approximately 200000 EUR annually for a one-directional MT system with a Scandinavian language at one end.

Our primary focus will be the path from prototype in the research lab to customers' PCs, and we will use Norway as the basis for our discussion. We assume that the conditions for commercialization are more or less the same in the other Scandinavian countries.

2   Systems

In the following paragraphs we will refer to two technical "scenarios" when needed for the discussion:

-   The LOGON demonstrator as developed at present (Lønning et al, this conference)
-   The PONS project (Dyvik 1995)

LOGON is described in another paper at this workshop. There are certain obstacles with the technological infrastructure chosen in LOGON which could be problematic for commercialization. The system makes use of the XLE platform for source language analysis, and XLE is owned by Palo Alto Research Centre (formerly Xerox PARC). The transfer and generation components are in LKB, which is licensed as open source software, we have been told. And finally, LOGON makes use of a bilingual dictionary owned by the Norwegian publisher Kunnskapsforlaget.

PONS is based on InterLisp, and the developers (Helge Dyvik and his group at the University of Bergen) have the relevant intellectual property rights (IPRs), see Dyvik (1995) for an introduction to the system.

In the discussion below we assume that the MT prototype systems have passed a first "suitability" check in the sense that they are regarded as highly promising by the academic research community.

3   Business Models

Generally, academic research groups are not, and perhaps should not, be concerned with business models for research prototypes. But if developers are interested in pursuing their ideas into products, there are at least two paths available when a prototype is ready:

- Include the prototype in the portfolio of some existing company - *the licensing model*
- The developers create a company on their own, and eventually include venture partner(s), if possible – *the new company model*

We will discuss each of them below, but before that we describe some aspects of "technology transfer offices".

3.1   Technology Transfer

Most Norwegian universities have some kind of "Technology Transfer Office" (henceforth TTO), whose primary goal is to commercialize ideas from their researchers. TTOs have the IPRs on behalf of the university's employees, and consequently the right to commercialize results which are developed inside the institution. This is a new arrangement in Norway, but it is quite normal in other countries. At NTNU in Trondheim the TTO is an ordinary limited company ("aksjeselskap"), owned by the university. Other Norwegian universities have slightly different models, but the University of Oslo seems to have decided to organize technology transfer more or less as NTNU does. The TTO will always let commercial concerns decide which ideas are to be pursued. Consequently, the TTO must use standard criteria in their project selection, but, crucially, a TTO should at least in theory have better qualifications to assess the technological potential by using the research staff at the hosting university.

In the years to come we assume that TTOs will play a key role in commercialization efforts of prototypes (including language technology and machine translation) arising from Scandinavian universities. Some initial questions in the commercialization process are

- Where and who are the customers? (domestic / abroad, general consumer market / business market, ...)
- Are the customers willing to pay for the product? If so, how much?
- Is it possible to estimate market size? Current or emerging competitors?
- What are the channels into the market?
- How should the product be presented to customers (stand-alone application, inclusion in office suites, etc)?
- How much work remains on application embedding (e.g. for making it a stand-alone Windows application) and who is expected to do it?

Compensation for developers at NTNU's TTO is only available after the application has generated profit.

Generally, the following conditions should be attested prior to serious commercialization attempts:
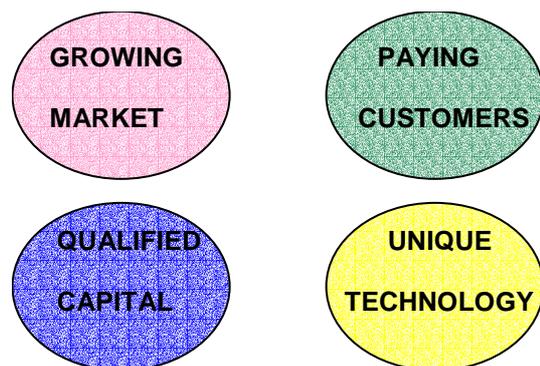

Figure 1: Commercialization Conditions (borrowed from Karl Klingsheim, TTO NTNU)

By "qualified capital" we mean licensees (see below), venture companies, private investors, "seed funding" services, and TTOs. Growth tendencies and willingness to pay among customers can, partly, be modelled by studying annual reports and technology analyses on the international scene, for instance companies like SYSTRAN and preferably other companies selling MT systems, so that the TTO and other actors in the commercialization process will possibly be able to refer to an international business segment. "Unique technology" should, in our context, be interpreted as solutions which enable people and/or companies to perform significantly better.

3.2   The Licensing Model

In the licensing model the TTO tries to license the prototype to some other company, which henceforth is called the licensee. The licensee will typically be a company with which the TTO has

more or less formalized cooperation. We assume that the prototype is made available for the licensee without any accompanying in depth market investigation performed by the TTO.

It is important for the licensee to decide whether the product is intended for private consumers or the enterprise market. The decisions made here are directly relevant for channel considerations, i.e. how to approach the customers. If the licensee has channels to the private consumer market, e.g. via chain stores selling electronics and software, then the licensee will have to make sure that the MT product is regarded as interesting by the managers of those stores. Based on experience, these channels are quite demanding in the sense that they require a turnover that exceeds reasonable levels from their point of view.

As mentioned earlier the market estimates are rough, and we have not yet worked out the costs for final application development, which is a technical issue about how the MT system should be finalized: As an application embedded in the licensee's existing products, as a stand-alone application, or as a third party product to some office suite collection, where Microsoft Office is the most relevant candidate. Given that we assume a market size at roughly 200.000 EUR these costs must be small. This leads us to claim that the MT prototype must be very close to a final product to persuade the licensee to bring it to the Scandinavian market.

What this means, in practice, is that the prototype must be delivered as an almost complete product, preferably with well-documented APIs which are compatible with APIs of major office suites. This job could be done by the TTO prior to a licensing agreement, but the TTO would only do so if it were fairly sure about the prototype's commercial potential.

We will not enter a discussion of other market segments, but think that the arguments above will apply there, too. To conclude this section, we think that a licensee would only be willing to bring a final, or nearly final, product to the market.

3.3 New Company

If it turns out to be impossible to find a licensee for the prototype a brand new company is another solution. This option requires that central developers take part in the creation of the new company if its primary product is the MT system. This is the most usual commercialization avenue from university labs, and the new company typically needs financial help from seed funding services, private investors, and a TTO or equivalent, but a substantial amount of strategic and technological work must be done by one or more of the developers.

In any case, the same requirements as laid out in figure 1 are still valid for a successful story. If seed funding services, private investors etc. are not convinced by the business plan for the company, the researchers will have to do all commercialization themselves.

Suppose that a small MT company is created. If it turns out that it is able to identify paying customers, generate a positive cash flow and survive on its own, it might be in a position for expansion, either by extending the market size with help from established vendors, or by extending the product portfolio by creating solutions for new language pairs, preferably in foreign markets. If relevant market analyses are positive, a venture company could be interested and willing to buy shares in the company, thereby expanding the financial basis. This is a critical phase for the company founders because venture capitalists will typically try to gain control over the company. Alternatively, the technology could be sold or licensed as in the previous model.

4   Unique Technology and Growing Market

Let us return to the factors "growing market" and "unique technology" from figure 1. In our context the relevant interpretation of "unique technology" is a solution which, in their own view, enables people and/or companies to do significantly better than previously in important areas. Technical MT internal details are for them quite irrelevant. The most important feature is a successful integration in their existing software solutions, which on the Scandinavian market are the key products from Microsoft. If users are forced into some unknown environments, not to mention into new operating systems or even new hardware, the marketing job becomes much harder. In effect, this means that on the private consumer market the MT product is more or less forced to integrate with Microsoft's translation API, or, alternatively, new interfaces must be developed and integrated in the various office components.

The "growing market" dimension is most important when venture companies or other investors are invited to take part in an extension phase of a young MT company. They will typically start with a close look at the company's annual reports over the last years and try to assess growth tendencies, both with respect to sold items and earnings. It is reasonable to assume that the figures they find will not be very exciting for a one way MT system from a Scandinavian language to for instance English (or vice versa), given that our turnover estimates in the introduction of this paper

are fairly trustworthy indications. The next and obvious question for involvement in the company is whether the market is growing, or has growth potential, for instance by identifying new segments or by expansion to other countries. If an export strategy is chosen, important cross lingual scalability questions of the MT system arise, and business strategies towards these markets become another question.

Given that there are affirmative answers to these issues, relevant questions for prototypes like LOGON and PONS would be how they scale up wrt new language pairs (grammar development, transfer rules, lexicon creation or acquisition, parser and generator behaviour, etc.), IPR issues (e.g. XLE, LKB, training corpora, dictionaries), and so on.

5   LOGON vs PONS

Copyright issues are very important for the licensing conditions. If the prototype has copyright holders who claim cuts in earnings or any other kind of copyright reimbursement, the commercialization efforts are likely to break down at TTO already. This could be a serious problem for the LOGON prototype because it does make use of XLE, and even more problematic if bilingual lexical resources from Kunnskapsforlaget are utilized. The PONS system is more interesting for commercialization in this respect because it does not have any copyright holders apart from the developer, but PONS does not have a realistic linguistic coverage, as it stands now. This tension between "translational" coverage and IPR claims is perhaps the most challenging aspect of an academic high quality MT system on its way towards a consumer product.

6   Conclusion

If ongoing academic MT research projects in Scandinavia have commercialization ambitions, we believe that proper qualifications at TTOs are very important. TTOs can support or perform market analyses and help formulate business plans which are anchored in reality.

We have made a rough estimate of the Scandinavian MT market by using Systran's official annual report. If this estimate is reliable we believe that the most realistic commercialization strategy of a Scandinavian MT prototype is through creation of a new language technology company, alternatively by allowing an existing software company to include the prototype for a nominal price. A licensing strategy, we have argued, is only possible if the prototype has been developed further into a mature application because of the relatively small market size.

References

Dyvik, Helge (1995). Exploiting Structural Similarities in Machine Translation. Computers and the Humanities 28:225 - 234 (1995).

Lønning, Jan Tore & al (2004, this conference). LOGON. Rasmat 2004, Uppsala, Sweden.